

I claim:

1. A method for reading addresses in more than one language, comprising the steps of:

- reading address characters using OCR means, said OCR means being directed to an anticipated language of said characters;
- depicting results of said reading in language-neutral transliteration form;
- determining and classifying address elements according to anticipated language related syntax rules, said address elements comprising said address characters; and
- verifying if each of said elements substantially match a database entry, said match comprising a defined degree of similarity, and said database comprising entries of acceptable read address elements with different, language dependent, transliteration variations.

20 2. The method according to claim 1, further comprising the steps of:

- prior to said step of reading address characters, recording an image of an address bearing surface;
- determining in said image regions comprising said address blocks, said step of determining in said image being performed by means of language related layout models, said models being generated from learning samples; and
- pictorially segmenting said address blocks so as to produce segmented image data.

3. The method according to claim 2, further comprising the steps of:

- feeding said segmented image data into a language decision unit;

- determining a corresponding language by comparing said blocks with language-typical feature sets, whereby said language has a highest comparison rate; and

5 - assigning said language as said anticipated language.

4. The method according to claim 3, further comprising the steps of: repeating said step of determining a corresponding language and assigning said language if said step of reading address characters fails with a previously assigned language.

10

15 5. The method according to claim 1, wherein if said step of reading address characters fails to resolve said address characters with said OCR means, reading identified words of said address in a word recognition unit, said word recognition unit comprising decision logic according to said anticipated language, and verifying results of said word recognition unit with said database.

20

25 6. The method according to claim 1, further comprising the steps of: repeating said steps of reading address characters, depicting results, determining and classifying address character elements with other languages than said anticipated language if said elements do not substantially correspond to database entries.

30

35 7. The method according to claim 4, further comprising the steps of: repeating said steps of reading address characters, depicting results, determining and classifying address character elements with other languages than said

anticipated language if said elements do not substantially correspond to database entries.

8. The method according to claim 1, wherein if said element substantially but not completely matches a database entry, changing said element to completely match said database entry.
5
9. The method according to claim 1, wherein at least one of said languages is non-Latin based.
10
10. A system for reading addresses in more than one language, comprising:
15
 - an optical character recognition (OCR) unit directed to anticipated languages of characters of said addresses, said characters being positioned in address blocks, said OCR unit comprising means for reading said addresses and depicting results in a language-neutral transliteration representation;
 - an address analysis unit for evaluating characters read by said OCR unit, said address analysis unit comprising means for determining and classifying address elements by reference to anticipated language-related syntax rules; and
20
 - an address interpretation unit for verifying identified address elements using an address database, said database comprising different, language-dependent transliteration variants for each database entry, said address being verified or accepted when each of said address elements is substantially similar to a database entry, wherein a level of similarity is predefined.
 - 25
- 30
- 35 11. The system according to claim 10, further comprising:

- means for generating an image of a surface containing address blocks;
- means for determining said address blocks based upon anticipated language related layout models,
- 5 said models generated from learning samples; and
- means for pictorially segmenting said address blocks.

12. The system according to claim 11, further
10 comprising a language decision unit, said language decision unit comprising:

- means for receiving said segmented image data; and
- means for designating an anticipated language by
15 comparing said blocks with language typical feature sets such that said anticipated language is a language having a highest degree of comparison with said blocks.

20 13. The system according to claim 12, further comprising a word recognition unit for reading parts of said address, said parts comprising words, said word recognition unit operable when reading results of said OCR unit are not verifiable, said word recognition unit
25 comprising decision logic of each anticipated language, and said word recognition unit further comprising means for feeding results to said address interpretation unit.